DOCUMENT RESUME

ED 334 228 TM 016 728

AUTHOR Newby, Tim; And Others
TITLE Instructional Diagnosis: Effective Open-Ended Faculty Evaluation.
PUB DATE Apr 91
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Classification; College Faculty; *College Students; Course Evaluation; *Evaluation Methods; *Faculty Evaluation; Higher Education; Instructional Effectiveness; Student Attitudes; *Student Evaluation of Teacher Performance; *Teacher Characteristics
IDENTIFIERS *Open Ended Questions; *Small Group Instructional Diagnosis

ABSTRACT

A faculty evaluation technique was used to classify responses from open-ended evaluations to determine elements of university teaching that were of greatest concern to students and to determine the feasibility of comparing subgroups by instructor gender, class size, and school. Small Group Instructional Diagnosis (SGID) was used--a form of course evaluation that offers students the opportunity to generate answers to open-ended questions. The comments from 147 SGIDs conducted between April 6, 1989 and May 1, 1990 at Purdue University (Lafayette) were compiled. The evaluations were obtained from 3,566 students and resulted in 2,422 comments that were classified by three raters into the following categories: (1) instructor; (2) learner interaction; (3) media; (4) content; (5) printed material; (6) grading; and (7) course policy. The instructor's personal characteristics emerged as the greatest concern to students, followed by the learner being actively involved in the learning process, and then by course content. The results show that responses can be: classified to identify elements of concern to students, and grouped to make comparisons by subgroup feasible. Such comparisons could be used to improve instruction. Sixteen figures and two tables are provided. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED334228

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

MARTI SHERMAN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Instructional Diagnosis:

Effective Open-ended Faculty Evaluation

Tim Newby

Marti Sherman

Sara Jane Coffman

Purdue University

Paper presented at the meeting of the American Educational Research Association
April 3-7, 1991, Chicago, IL

TM016728

BEST COPY AVAILABLE

Instructional Diagnosis:
Effective Open-ended Faculty Evaluation

Debates concerning the value, purpose, appropriate procedures, and effectiveness of course evaluation at the university level have received a great deal of attention in professional journals. The purpose of this study is not to add to these debates, but to take a fresh look at open-ended evaluation. Open-ended evaluation is rarely used for research purposes because there is no effective procedure for comparison. It does seem, however, that much could be learned from analyzing the results of a number of open-ended evaluations.

Four issues will be addressed in this section — students as evaluators, midterm evaluations and feedback, rating scales for evaluation of university courses, and open-ended evaluations of university courses.

Students as Evaluators In a lengthy comparison study of evaluations by students, colleagues, administrators, external evaluators, and instructor self-evaluations on the same course, Feldman (1989) found the highest relative similarity in evaluation (as indicated by the product-moment correlation among ratings made by the different sources of evaluation) was between students and colleagues and between students and external evaluators. The next highest similarities were between administrators and colleagues and between students and administrators. The lowest similarities were between instructor self-evaluations and students, self and colleagues, and self and administrators. This comparison, based on more than 40 studies, supports other research (Centra, 1977; Cohen, 1981; Sullivan & Skanes, 1974) that confirm that students are reliable evaluators of university courses.

Midterm Evaluations and Feedback Overall and Marsh (1979) found that when midterm feedback was presented by an evaluator with a discussion of how the evaluations might be used to improve teaching effectiveness, final evaluations, examination performance, and subsequent course enrollment were significantly higher than when feedback was presented only in written form. Similarly, in a meta- analysis, Cohen (1980) found that feedback with consultation on midterm evaluations resulted in higher final ratings than feedback with no consultation. Abbott, Wulff, Nyquist, Ropp, and Hess (1990) found that students preferred having their opinions sought at midterm. In addition, their findings indicate that students were more satisfied when the instructor offered an extended and specific reaction to their evaluations. Midterm evaluations offer the opportunity to make adjustments during a course such that the effectiveness of these adjustments can be evaluated at the end of the course.

Rating Scales Abrami (1989) notes that few universities in North America evaluate teaching effectiveness without the use of student evaluations, and yet, there has been little agreement on uniform evaluation procedures. The most common form of evaluation is Likert rating scales using predetermined statements. The effectiveness of rating scales has been the topic of considerable research (Abrami, 1988, 1989; Johnson, 1989; Marsh, 1987). Abrami (1989) argues that the dimensions of effective teaching have not been sufficiently established, therefore, it is inappropriate to use specific items as a means of comparing instructors . For example, "The instructor encourages students to participate in class discussion" does not measure a dimension of effective teaching equally for a class of 20 as for a class of 150 where participation in class discussion is less likely to occur. He does, however, support the use of rating scales for more global items such as "How would you rate this instructor in overall ability?"

Rating scales such as the "cafeteria" form in which the instructor selects the items to be used for evaluation raise additional questions. The assumption is that instructors know the qualities of good teaching and are objective in selecting items that effectively measure their own strengths and weaknesses. However, Feldman's (1989) review indicated that instructors' self-evaluations were least similar to those of students, colleagues, or administrators. This would suggest that instructors are not the most objective evaluators of their own instruction nor of the items on which they should be evaluated. This does not imply that rating scales such as the cafeteria form should not be used for evaluation. It does raise questions, however, about the selection of the items and the use of the evaluation. Consistent with Abrami (1989), The Center for Instructional Services at Purdue University recommends, "If CAFETERIA ratings are to be used in promotion and tenure decisions, additional precautions should be taken to insure that data are obtained properly and interpreted fairly and intelligently." The Center offers a number of guidelines, including, "Use only the five University core items to make comparisons between classes or between instructors.."

Open-ended Evaluations Compared to rating scales, open-ended evaluations have been used primarily for the purpose of instructional improvement rather than for comparison. Tiberius et al. (1989) demonstrated that while formative evaluations using standardized rating scales can improve teaching, improvements are enhanced and sustained by supplementary feedback methods. They used small group discussions with students by an outside facilitator as feedback and found that while the use of rating scales produced improvement, it was not sustained into the subsequent semester. The use of feedback from group discussion in addition to rating scales produced greater improvements in teaching performance; moreover, it was sustained into the next semester. Abbott et al. (1990) found that students were more satisfied with open-ended evaluations than rating scales when administered at midterm. There was little difference in their preference when the evaluation was administered at the end of the course.

Open-ended evaluations provide more specific information and are more effective for the purpose of instructional improvement, but they are not easily quantified for purposes of comparison. Rating scales are quantifiable and easily used for comparison, but there is little agreement on the items that measure effective teaching, especially across a variety of teaching situations. Is it feasible to adopt a methodology of evaluation that uses both forms of evaluation in conjunction with each other?

The objectives of this study were to classify the responses from open-ended evaluations to determine the elements of university teaching that are of greatest concern to students and to determine the feasibility of comparing subgroups by instructor gender, class size, and schools.

EVALUATION PROCEDURE

Small Group Instructional Diagnosis (SGID) is a form of course evaluation developed at the University of Washington and offered by the Center for Instructional Services at Purdue University. This evaluation technique, administered by trained facilitators, offers students the opportunity to generate answers to open-ended questions. There are currently six trained facilitators at Purdue University. The procedures for conducting a Small Group Diagnosis are very precise and consistently administered.

The facilitator is introduced to the class by the instructor. The instructor then leaves. The facilitator divides the class into small groups of five to seven students. Each small group is given a form containing three questions: "What do you like about this course? What do you think needs improvement in this course? What specific suggestions do you

have for changing this course?" The small groups are given six minutes to respond to the three questions. There must be group consensus for each response to be included.

After six minutes, the facilitator calls the class back together. Each group is asked, in turn, to state one of their responses. The responses are written on the chalkboard (or similar display method). This continues until all responses have been stated. Everyone must then agree to all of the responses. If there is disagreement, the response is reworded or dropped. This procedure takes approximately 15 - 20 minutes of class time. The facilitator later meets with the instructor to discuss the results and clarify any questions about the responses.

Small Group Instructional Diagnosis provides an opportunity to ascertain the elements of a course that are of greatest concern to the students. The procedure was carefully designed to limit the amount of time for responding to the questions so that it can tap the most significant concerns. By requiring group consensus, it is not possible for one or two students to skew the results.

METHOD

To determine the elements of university teaching that are of greatest concern to students, the comments from 147 Small Group Instructional Diagnoses conducted between April 6, 1989 and May 1, 1990 at Purdue University were compiled. These evaluations involved 3,566 students. Three raters were asked to identify five to seven categories into which the 2,422 comments could be classified. Five categories were common to all three raters. A discussion among the raters to determine the resolution of the ones that were not held in common resulted in agreement on the following seven categories.

Instructor - All statements pertaining to the personal characteristics of instructors or the instructor's presentation of the material to the class. This category includes statements about the instructors' personality, attitude, demeanor, presence, manner, commitment, style, favoritism, rapport, availability, presentation style, clarity of instructions and answers to questions, examples, humor, classroom atmosphere, use of class time, making proper assumptions about students.

Learner Interaction - All comments pertaining to the learner being actively involved in the learning process. This category includes statements about group discussions, questions, hands-on activities, practical experiences, assignments, projects, homework, amount of work load, due dates, feedback.

Media - All comments pertaining to films, videos, guest speakers, field trips. Comments involving a trip that resulted in hands-on experience for the students were classified under Learner Interaction.

Content - All comments pertaining to the selected content of the course. The overall content may or may not be within the control of the instructor. This category includes statements about the content's relevance, practicability, applicability, appropriateness of level of content.

Printed Material - All comments pertaining to printed materials used by the class. This category includes statements about books, handouts, readings, printed notes, materials on reserve in the library, and software used like a textbook.

Grading - All comments pertaining to earning a grade. This category includes statements about mastery learning, grading philosophy, tests, exams, quizzes, and help sessions.

Course Policy - All comments pertaining to decisions about the course itself. These decisions may or may not be controlled by the instructor. This category includes statements about equipment and facilities, class size, prerequisites, meeting time of class, team teaching, teaching assistants, lab technicians, substitutes.

Three raters independently classified the statements into the seven categories. Agreement of two of the three raters constituted acceptance of the classification. All comments not receiving agreement were discussed and agreement was reached by the raters. Generally, when agreement was not reached initially, it was because the statement was not understood by the rater. For example, comments like "We like the swine camp" or "We suggest more vignettes" were difficult to classify. The individual who conducted the Small Group Diagnosis was consulted for clarification, and the statement was then classified and agreed upon by the raters.

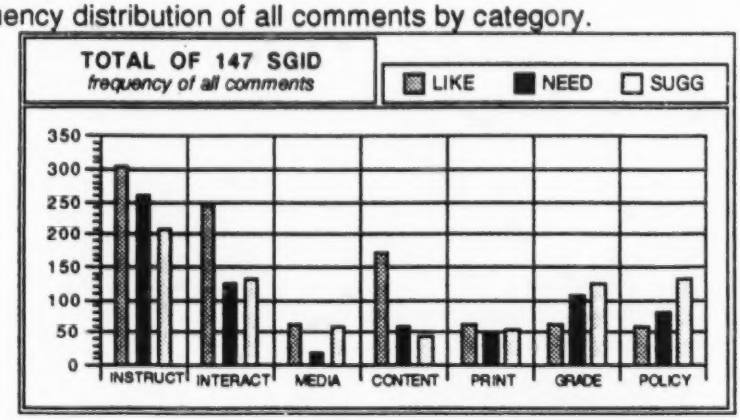
RESULTS

The results of this investigation are descriptive in nature. The comments on each of the three questions and in each of the seven categories were tallied. The results were then graphed for the TOTAL, by INSTRUCTOR GENDER, by CLASS SIZE, and by SCHOOLS.

Total

Overall, results indicate that the greatest number of comments focus on the personal characteristics of the instructor followed by comments of things the students liked concerning learner interaction. On grading and course policy, students tended to comment more frequently about needing improvement and suggestions for change rather than things they liked. The frequency of the categorized comments for the total of all of the Small Group Instructional Diagnoses (SGID) are presented in Figure 1. "LIKE, NEED, and SUGG" refer to the three questions: "What do you like about this course? What do you think needs improvement in this course? What specific suggestions do you have for changing this course?" The divisions listed across the bottom of the graph refer to the seven categories to which the comments refer: Instructor, Learner Interaction, Media, Course Content, Printed Material, Grading, and Course Policy.

Figure 1. Frequency distribution of all comments by category.



The Instructor category received the most frequent comments. Of the seven categories, 31.96% of all the comments referred to the personal characteristics of the instructor.

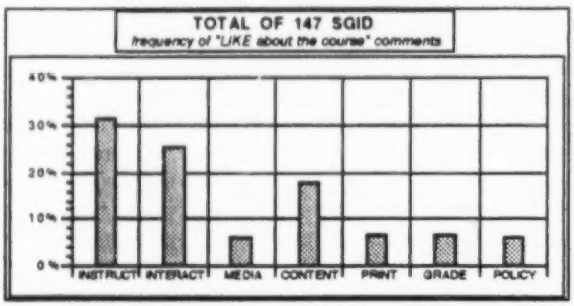
Learner Interaction, referring to the learner being actively involved in the learning process, received the second highest frequency of comments: 20.73%. It is interesting to note that most of these comments were in response to things they liked. This would suggest that when given the opportunity to be actively involved, it is recognized and appreciated by the students.

Grading accounted for 12.10% of all the comments. Comments concerning the need for improvement and suggestions for change outweighed comments of things they liked in this category. It was surprising that this category ranked third in overall comments and that it accounted for such a small percentage of the total comments.

Course Content (11.27%) and Course Policy (11.19%) ranked third and fourth respectively. The two categories that received the fewest number of comments were Printed Materials and Media with 6.94% and 5.82% of the overall comments respectively.

Of the total comments made in response to the three questions, 39.93% of the responses were to the question concerning things they like, 31.21% in response to the question about specific suggestions, and 28.86% in response to things that need improvement. Each of these questions was looked at separately. The comments were converted to a percentage by dividing the comments in each category by the total number of responses to each question. Figure 2 shows the division by category of only the responses to the question "What do you like about this course?"

Figure 2. Percentage of responses in each category to question pertaining to "LIKEs".

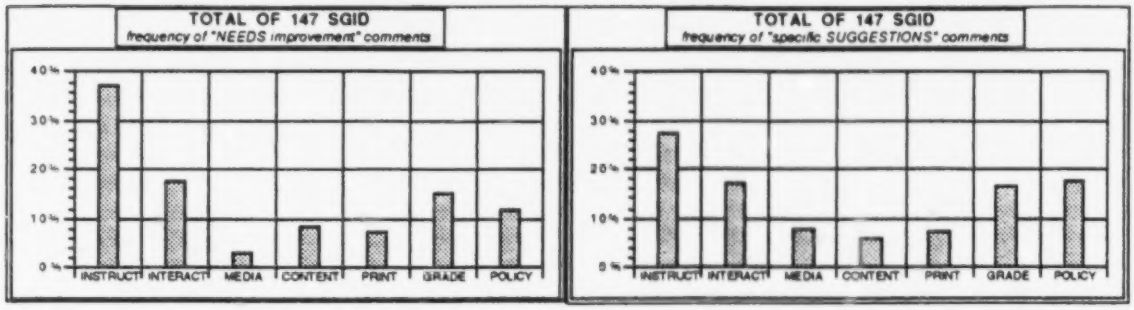


When discussing things they like about the course, the most frequently discussed category is Instructor followed by Learner Interaction. This agrees with the overall comments. Content ranked third for this question followed by Grading, Printed Materials, Media, and Course Policy.

Figures 3 and 4 look at the responses to the questions "What do you think needs improvement in this course?" and "What specific suggestions do you have for changing this course?" The same method of calculating the percentage for each category was used. Concerning things that need improvement, the category that received highest percentage of comments was Instructor followed by Learner Interaction and then Grading. These categories were followed in order by Course Policy, Content, Printed Material, and then Media.

Figure 3. Percentage of responses in each category to the question pertaining to "NEEDS Improvement".

Figure 4. Percentage of responses in each category to the question pertaining to "Specific SUGGestions for change".



In response to the question "What specific suggestions do you have for changing this course?" the category most frequently commented on is still Instructor, but Course Policy ranks second, Learner Interaction ranks third, and Grading ranks fourth. These are then followed in order by Media, Printed Materials, and Content.

It is interesting to note the shift in rank of frequency of comments in response to each question for the different categories in comparison to the overall ranking of each category. Table 1 compares these rankings for the total responses to all three questions collectively and according to frequency within each question.

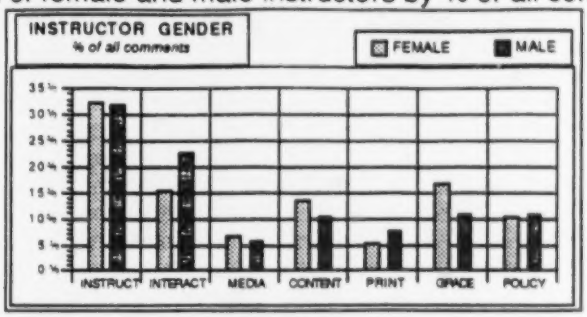
Table 1
Rank by Frequency of Comments in Each Category

RANK	TOTAL	LIKE	NEED	SUGG
1	Instructor	Instructor	Instructor	Instructor
2	Interaction	Interaction	Interaction	Policy
3	Grading	Content	Grading	Interaction
4	Content	Grading	Policy	Grading
5	Policy	Materials	Content	Media
6	Materials	Media	Materials	Materials
7	Media	Policy	Media	Content

Instructor Gender

The results were examined according to instructor gender. Of the 147 classes evaluated, 94 had male instructors and 38 had female instructors. The remaining fifteen were taught by more than one instructor and, therefore, were not included in this analysis. Due to the nature of the evaluation, unequal numbers of responses occur for each question and for each class that is evaluated. For the purpose of comparing groups, frequencies have been calculated as percentages of total comments or percentage of comments within a question for each group. Figure 5 compares the total of all comments for male and female instructors.

Figure 5. Comparison of female and male instructors by % of all comments.



The results indicate similarities for most categories. Learner Interaction received considerably more comments for male instructors (22.5%) than for female instructors (15.19%). Comments concerning Grading had the opposite relationship. For female instructors, 16.77% of all comments referred to Grading compared to 10.84% for male instructors.

Each of the three questions was analyzed separately with percentages calculated only for responses to that question.

Figure 6. Comparison of female and male instructors by % of responses to the question pertaining to "LIKES".

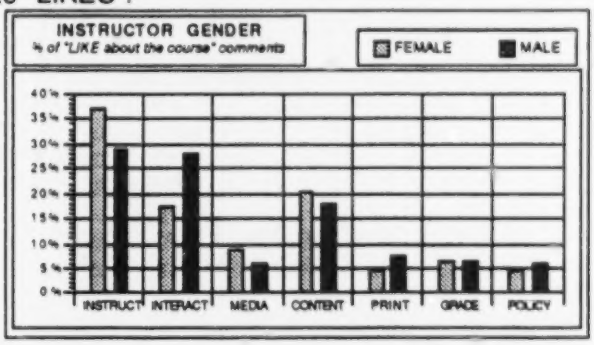


Figure 7. Comparison of female and male instructors by % of responses to the question pertaining to "NEEDS improvement".

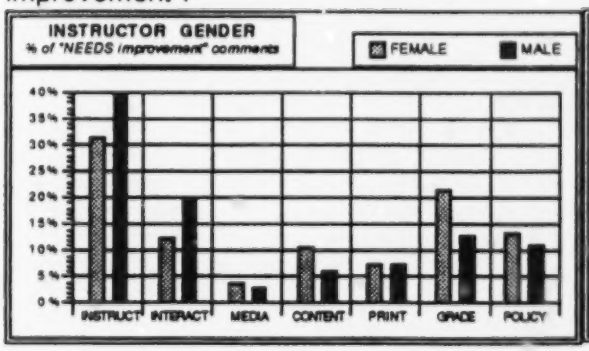


Figure 8. Comparison of female and male instructors by % of responses to the question pertaining to "specific SUGGestions".

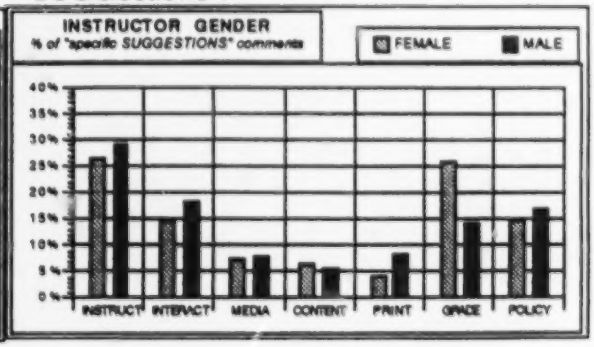


Figure 6 graphs the responses to the question "What do you like about this course?" For male instructors, there were a similar number of responses to this question concerning Instructor and Learner Interaction: 28.93% and 27.77% respectively. Students commented more frequently about personal characteristics of the instructor for female instructors than for male instructors when listing things they liked. However, the Learner Interaction category received more responses for male instructors than for female instructors.

Figure 7 represents the responses to the question "What do you think needs improvement in this course?" The largest gap in the percentage of comments about things that need improvement was in the Grading category: 21.28% for female instructors and 12.71% for male instructors. Again there was a large difference between the comments referring to the Instructor for male instructors (39.62%) and for female instructors (31.38%).

Comments concerning specific suggestions for change in the course are represented in Figure 8. The greatest difference between male and female instructors was in the category of Grading. Specific suggestions for changes in grading accounted for 25.93% of the responses for female instructors and for only 14.12% for male instructors.

It is interesting to note that in the last four Figures the relationship of male to female in the Learner Interaction category remains the same. In every case, males received more comments about Learner Interaction, positive and negative. In order to look at this more closely, the Learner Interaction category was calculated for the percentage of the total comments. For male instructors, 22.5% of the total comments were in the Learner Interaction category: 10.59% in response to Likes, 5.99% in response to Needs, and 5.92% in response to Suggestions. For female instructors, 15.19% of the total comments were in this category: 7.12% in response to Likes, 3.64% in response to Needs, and 4.43% in response to Suggestions.

The comparison of male and female instructors is best seen in Table 2 and Figure 9 where the distribution of all the comments in response to all three questions is displayed. Figure 10 represents the total of all comments for all of the evaluations that were analyzed. It is included here with Figure 9 so that the distributions for Male and Female Instructors can be compared to the total of all evaluations.

Table 2
Percentage of All Comments in Each Category

	FEMALE INSTRUCTORS				MALE INSTRUCTORS			
CATEGORY	Likes	Needs	Suggs	Total	Likes	Needs	Suggs	Total
Instructor	15.03%	9.34%	7.91%	32.28%	11.03%	11.78%	9.26%	32.07%
Interaction	7.12%	3.64%	4.43%	15.19%	10.59%	5.99%	5.92%	22.05%
Media	3.48%	1.11%	2.22%	6.80%	2.21%	0.82%	2.52%	5.55%
Content	8.23%	3.16%	1.90%	13.29%	6.81%	1.83%	1.76%	10.40%
Print Matter	1.90%	2.22%	1.27%	5.38%	2.77%	2.27%	2.71%	7.75%
Grading	2.69%	6.33%	7.75%	16.77%	2.52%	3.78%	4.54%	10.84%
Policy	1.90%	3.96%	4.43%	10.28%	2.21%	3.28%	5.42%	10.90%
Total	40.35%	29.75%	29.91%		38.12%	29.74%	32.14%	

Figure 9. Distribution of all comments for FEMALE and MALE instructors.

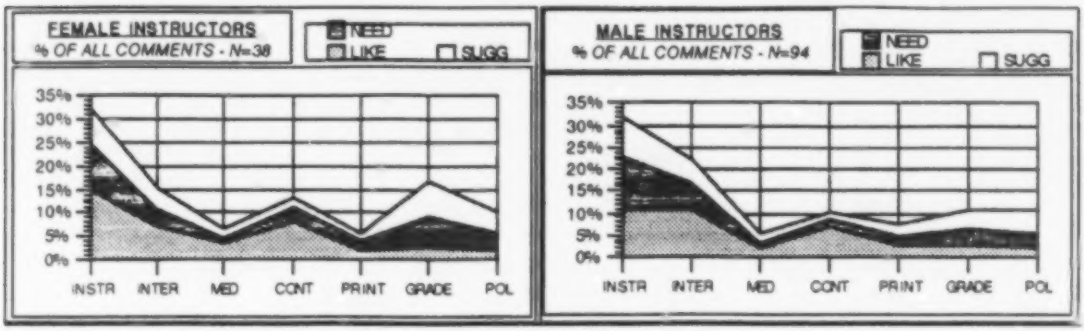
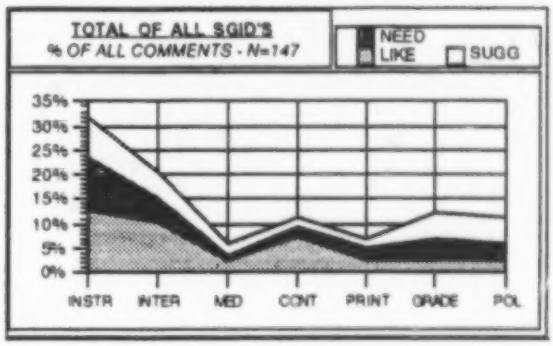


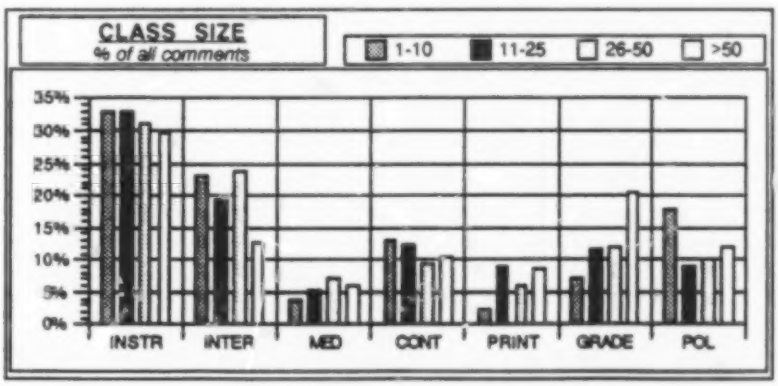
Figure 10. Distribution of all comments for ALL evaluations.



Class Size

The evaluations were divided into four class sizes: 1-10, 11-25, 26-50, and >50. Figure 11 shows the distribution of all of the comments. This was calculated by dividing the number of comments in each category by the total number of comments. There were 34 evaluations for Class Size 1-10, 53 for Class Size 11-25, 37 for Class Size 26-50, and 23 for Class Size >50.

Figure 11. Distribution of all comments for each category by class size.



BEST COPY AVAILABLE

As would be expected, the comments for Learner Interaction decline considerably for classes of >50. It is interesting to note the increase in the number of comments in this category for size 26-50 (19.3%) over size 11-25 (23.92%). Frequency of comments about Media increase according class size. Comments about Grading also increase with class size taking a dramatic jump in class sizes >50. Course Policy ranks third of all the categories for size 1-10, then drops to sixth for size 11-25 with a steady increase through the other class sizes.

In order to determine how these comments were distributed among the three questions, the responses to each question were calculated separately. Figures 12, 13, and 14 graph the distribution within questions.

Figure 12. Comparison of class size by responses to the question pertaining to "LIKES".

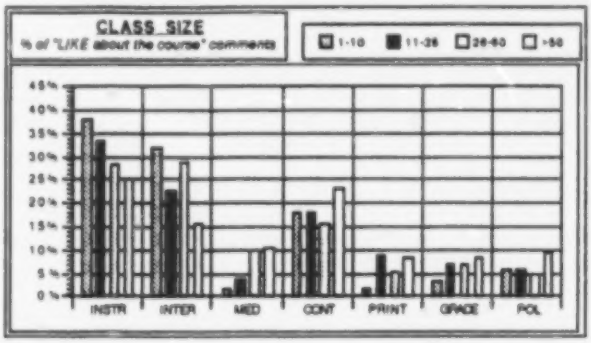
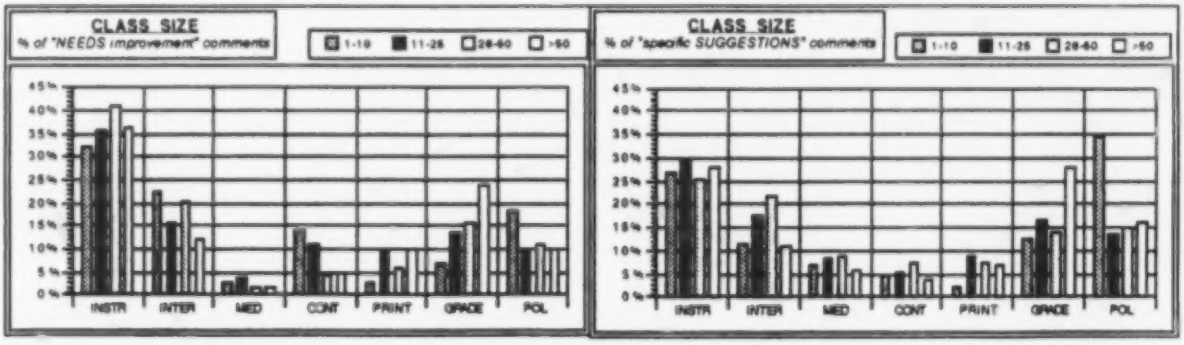


Figure 13. Comparison of class size by responses to the question pertaining to "NEEDS improvement".

Figure 14. Comparison of class size by responses to the question pertaining to "specific SUGGestions".



When discussing only the things they "LIKE", comments about the Instructor decline steadily with class size. For Class Size 25-50, comments about Learner Interaction outnumber those about Instructor. This is particularly interesting to note because in all of the subgroup analyses of this data, this occurs only one other time: in the School of Pharmacy, which will be examined later. Media receives increasing numbers of positive comments as the class size increases. Content, Grading, and Course Policy receive more "LIKE" comments for Class Size >50 than do the other three class sizes.

When responding to things that "NEED" improvement, comments about Grading increase with class size while comments about Content decrease with class size. Learner Interaction maintains about the same relative relationship that it had for "LIKE" comments.

In response to "SUGGestions" for change, there is an increase in comments about Media, Grading, and Course Policy and a decrease in comments about the Instructor. Class Size 1-10 has considerably more comments about Course Policy (34.45%) than for Instructor (26.89%). This, again, is particularly interesting because it is unusual for any category to receive more responses than the Instructor category.

Figure 15 shows the distribution of all the comments for each class size. Figure 10 is repeated here so that each Class Size subgroup can be compared to the total.

Figure 15. Distribution of all comments by CLASS SIZE.

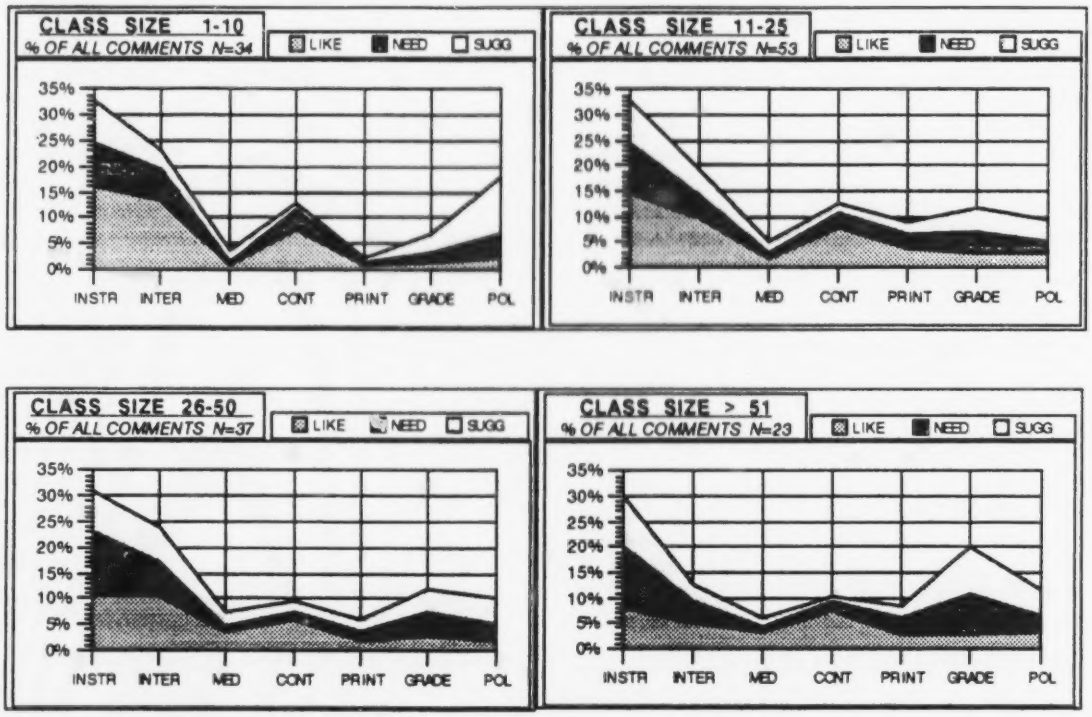
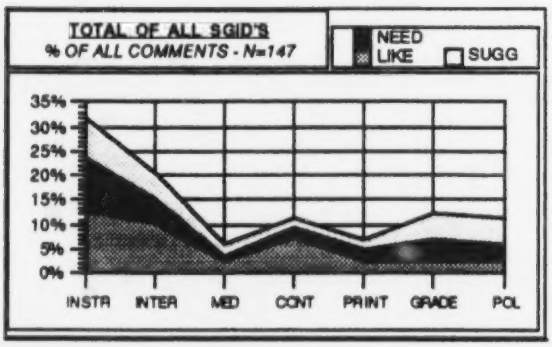


Figure 10. Distribution of all comments for ALL evaluations.



BEST COPY AVAILABLE

Schools

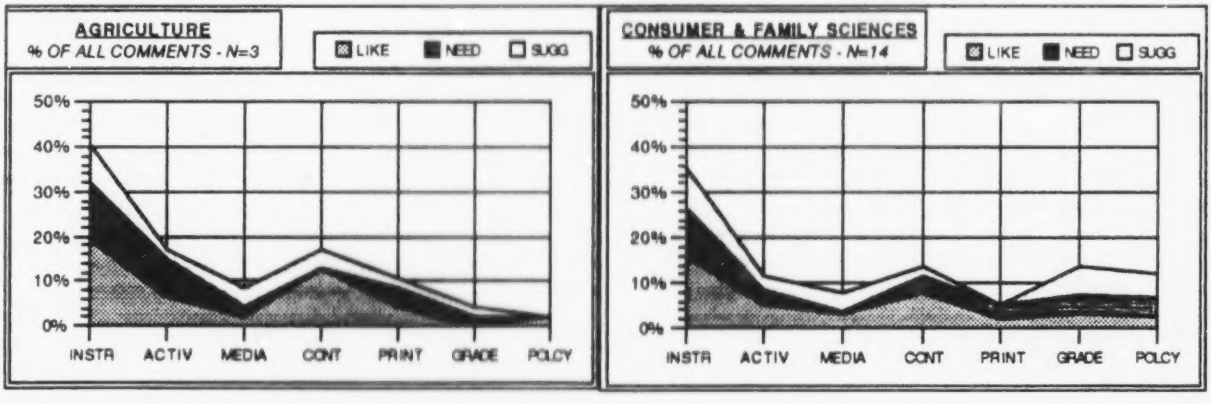
Ten schools were represented. The number of evaluations for each school varied as follows:

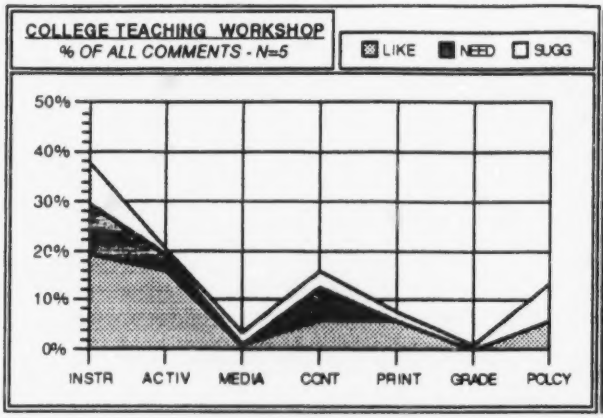
School of Agriculture	3
School of Consumer and Family Sciences	14
School of Education	9
School of Engineering	8
School of Liberal Arts	6
School of Nursing	8
School of Pharmacy	29
School of Science	1
School of Technology	47
School of Veterinary Medicine	17
College Teaching Workshop	5

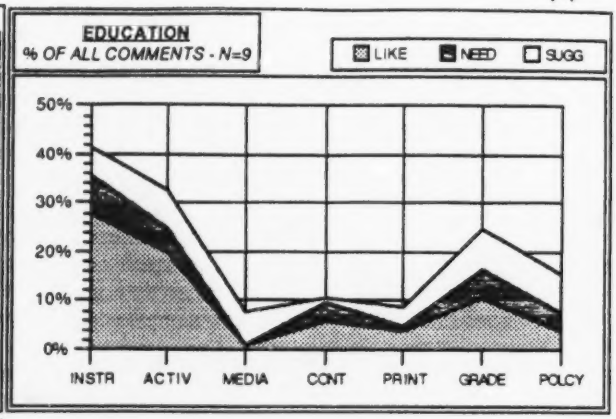
The Center for Instructional Services at Purdue University offers College Teaching Workshops throughout the year. Evaluations for these workshops are included even though they do not represent a school.

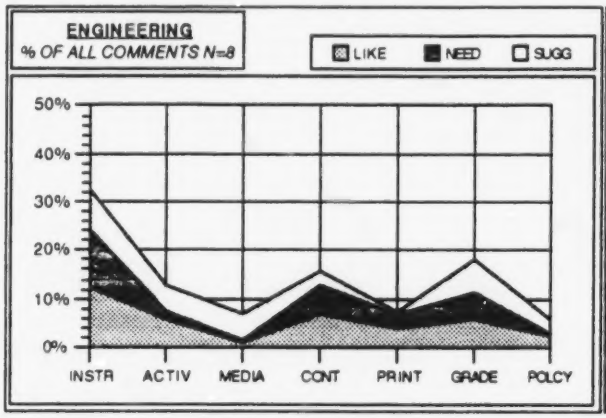
Figure 16 shows the distribution of all of the comments in response to each of the three questions for all of the schools and the College Teaching Workshops. Displaying them in this manner allows for comparison of comments among the various schools. Each school can also be compared to Figure 10 which represents the total for this data set.

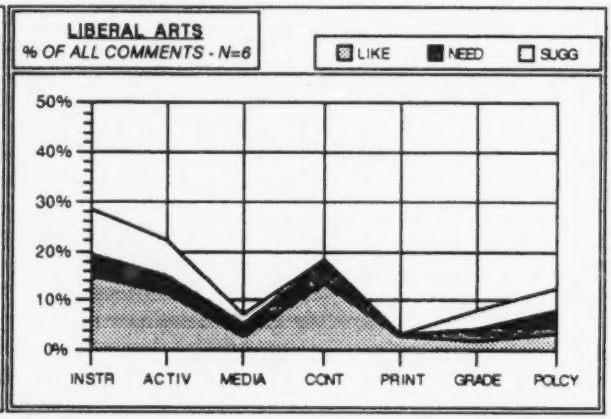
Figure 16. Distribution of all comments by schools.

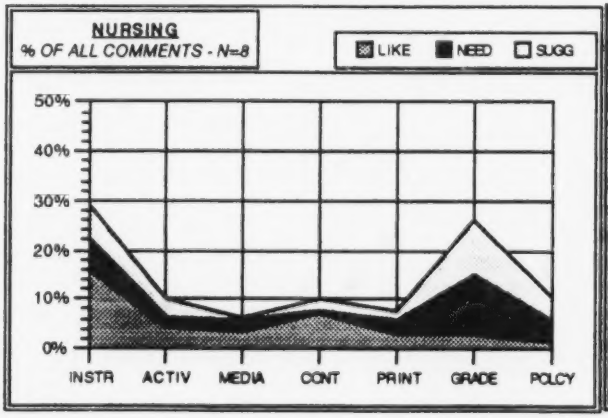


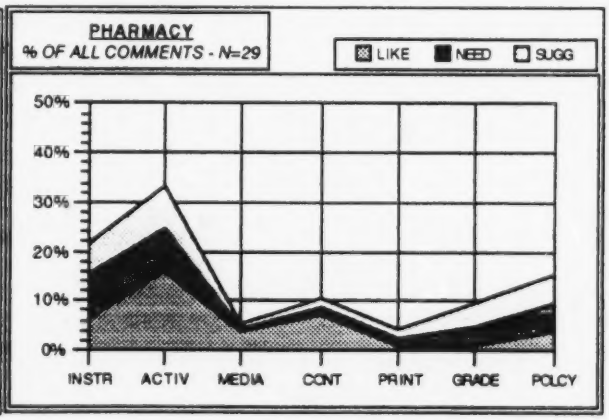




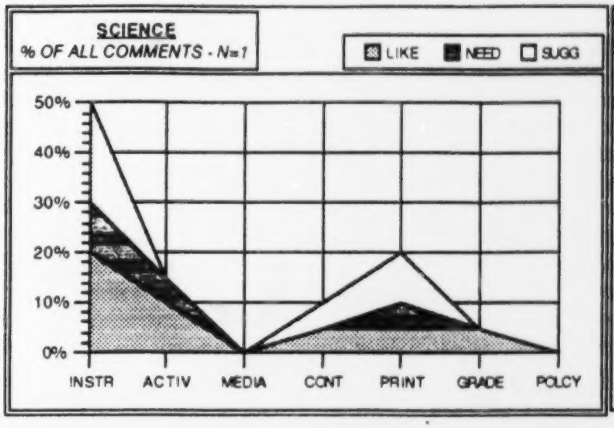


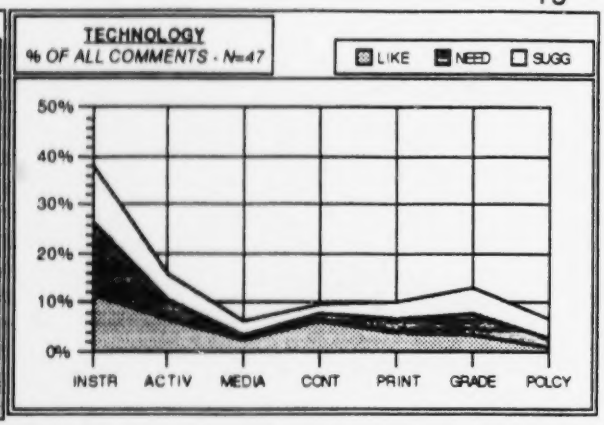


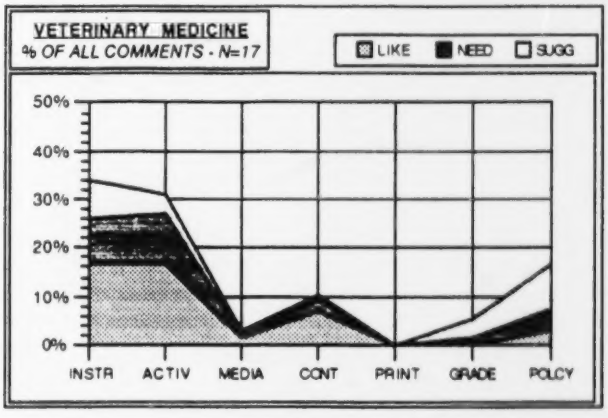




BEST COPY AVAILABLE







DISCUSSION

The results of this study present a fresh look at open-ended evaluation and offer interesting possibilities for further study. The first objective of this study was to develop a technique for classifying the responses to open-ended course evaluations using Small Group Instructional Diagnosis. It was demonstrated that a large number of responses could be placed into categories. Using this procedure, it is possible to identify the elements of teaching that are of greatest concern to university students. The instructor's personal characteristics emerged as the greatest concern to the students, followed by the learner being actively involved in the learning process and then by course content. From this sample, we can conclude that students have far greater concern about what they are taught, that they are involved with their learning, and the manner in which they are taught than they are about grading procedures, course policies, the use of media, or printed material. This suggests a greater interest in the true "meat" of the educational process than one might expect. This is even more compelling when you consider that 86% of these evaluations were from undergraduate classes. In order to more clearly identify which personal characteristics of the instructor are of greatest concern, it would be interesting to further break down the comments within the instructor category.

The second objective was to determine the feasibility of comparing subgroups such as instructor gender, class size, and schools. The data was graphed in a manner which allowed for comparison. For descriptive purposes, this method offers interesting possibilities. When the information from a single evaluation is graphed in the same

BEST COPY AVAILABLE

manner, it can be compared to the subgroup(s) of interest to that instructor. For example, comparisons can be made within a school for the purpose of improving instruction for that group of instructors.

This study leads to suggestions for future research. A study that is currently being conducted will look at the correlation between Small Group Instructional Diagnosis and final course evaluations using a rating scale. Investigation into improvements in final course evaluations after using the Small Groups Instructional Diagnosis would be of interest. The feasibility of schools or departments using this form of evaluation to determine and then evaluate instructional goals could to be investigated. Additional studies with larger samples could provide greater insight into the dimensions of effective teaching.

REFERENCES

Abbott, R. D., Wulff, D. H., Nyquist, J. D., Ropp, V. A., & Hess, C. A. (1990). Satisfaction with processes of collecting student opinions about instruction: The student perspective. Journal of Educational Psychology, 82, 201-206.

Abrami, P. C. (1988). SEEQ and ye shall find: A review of Marsh's "Students' evaluation of university teaching". Instructional Evaluation, 9(2), 19-27.

Abrami, P. C. (1989). How should we use student ratings to evaluate teaching? Research in Higher Education, 30, 221-227.

Centra, J. A. (1977). Student ratings of instruction and their relationship to student learning. American Education Research Journal, 14, 17-24.

Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. Research in Higher Education, 13, 321-341.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. Review of Educational Research, 51, 281-309.

Feldman, K. A. (1989). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues. administrators, and external (neutral) observers. Research in Higher Education, 30, 137-194.

Johnson, G. R. (1989). Faculty evaluation: Do experts agree? The Journal of Staff, Program, & Organization Development, 7, 22-28.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. International Journal of Educational Research, 11, 253-388.

Overall, J. U., & Marsh, H. W. (1979). Midterm feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes. Journal of Educational Psychology, 71, 856-865.

Sullivan, A. M., & Skanes, G. R. (1974). Validity of student evaluations of teaching and the characteristics of successful instructors. Journal of Educational Psychology, 66, 584-590.

Tiberius, R. G., Sackin, H. D., Slingerland, J. M., Jubas, K., Bell, M., & Matlow, A. (1989). The influence of student evaluative feedback on the improvement of clinical teaching. Journal of Higher Education, 60, 665-681.